Project – Forecasting Sales

04/2/2024
Tirthraj Bhatt
tb19950n@pace.edu
Practical Data Science
Data Science
Seidenberg School, Pace University

Agenda

- Executive summary
- Project plan recap
- Data
- Exploratory data analysis
- Modeling method
- Findings
- Recommendation & technical next steps

Executive Summary

- The goal of this project is to develop a predictive model that can forecast the quantity of products for each category in our ecommerce inventory. By accurately predicting product quantities, we aim to optimize inventory management processes, mitigate the risk of overstocking or understocking, and ultimately improve operational efficiency and customer satisfaction.
- We will utilize time series forecasting models, such as ARIMA (AutoRegressive Integrated Moving Average), and machine learning algorithms like Random Forest Regression, to predict product quantities in ecommerce inventory data.

Project plan recap

Deliverable	Dues Date	Status
Data &EDA	03/19/2024	Completed
Method, Findings, & Recommendations	04/2/2024	Completed
Final Presentation	04/16/2024	Not Started

Data

➢Data Details

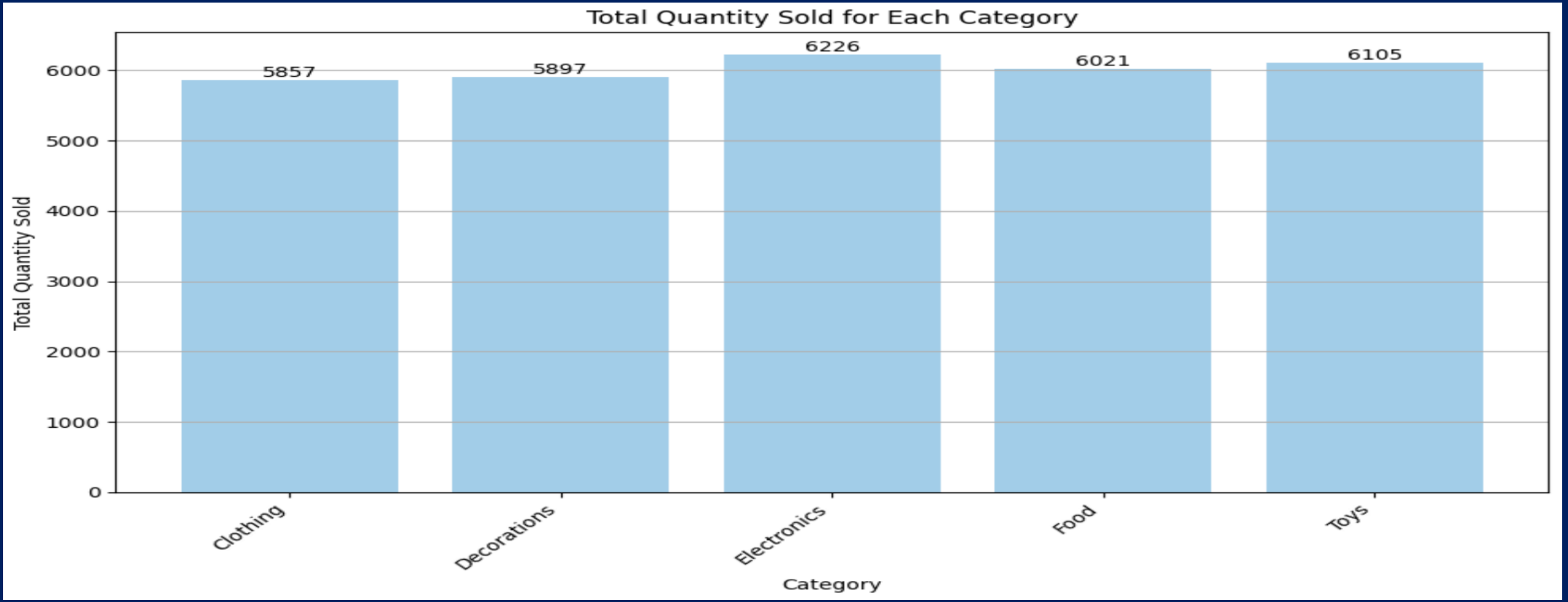
- Data Source
- Size - 10000 (rows) x 25 (Column)
- 12/2023
- The dataset comprises various features related to ecommerce transactions. After EDA, certain columns, including 'TransactionID', 'CustomerID', 'StoreID', 'OnlineOrderFlag', and 'DeliveryTime', were excluded due to their lack of correlation with the target variable, 'Quantity'.
- Additional features were derived from the 'Age' column, resulting in the creation of a new column representing the age range of customers.

Exploratory Data Analysis

EDA

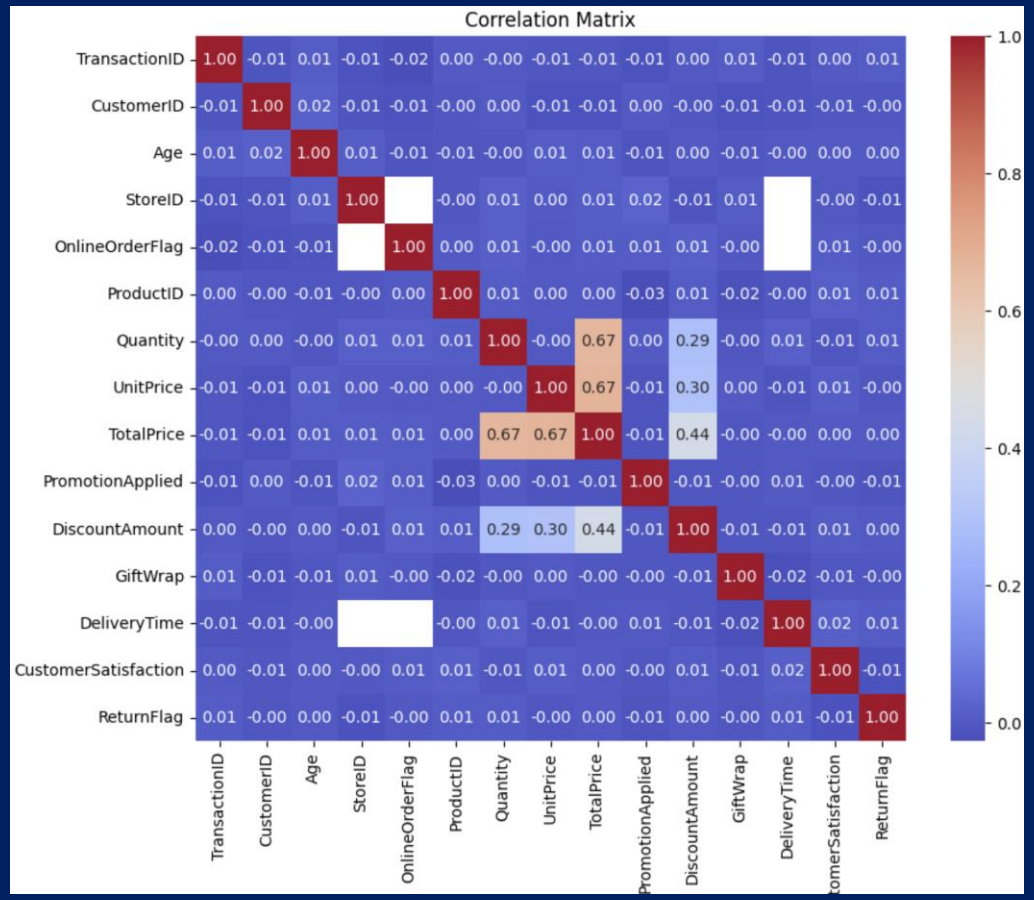
Examined the distribution of the target variable 'Quantity' and which clearly shows the trend of the quantity for different Years

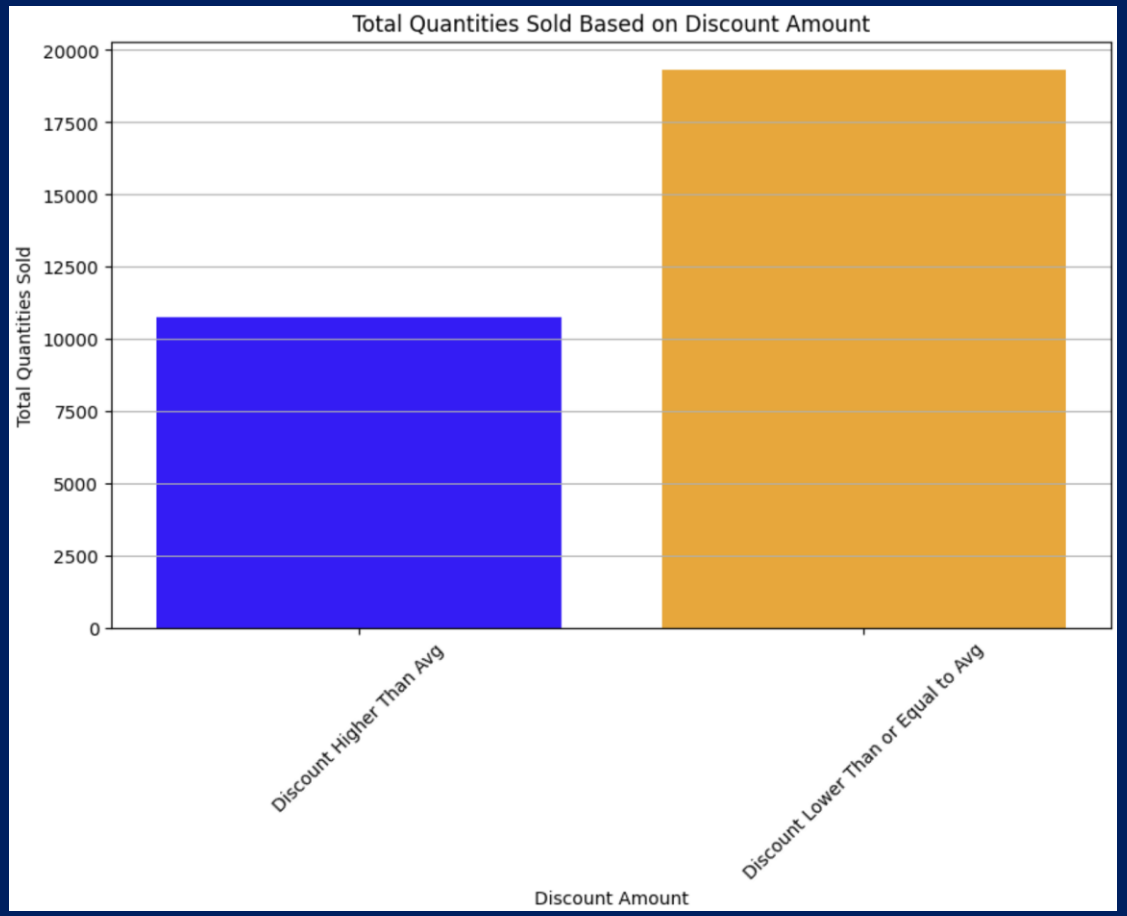
EDA



Examined the total quantity sold for each category and we can clearly see that Electronics is highest selling category followed by toys

EDA





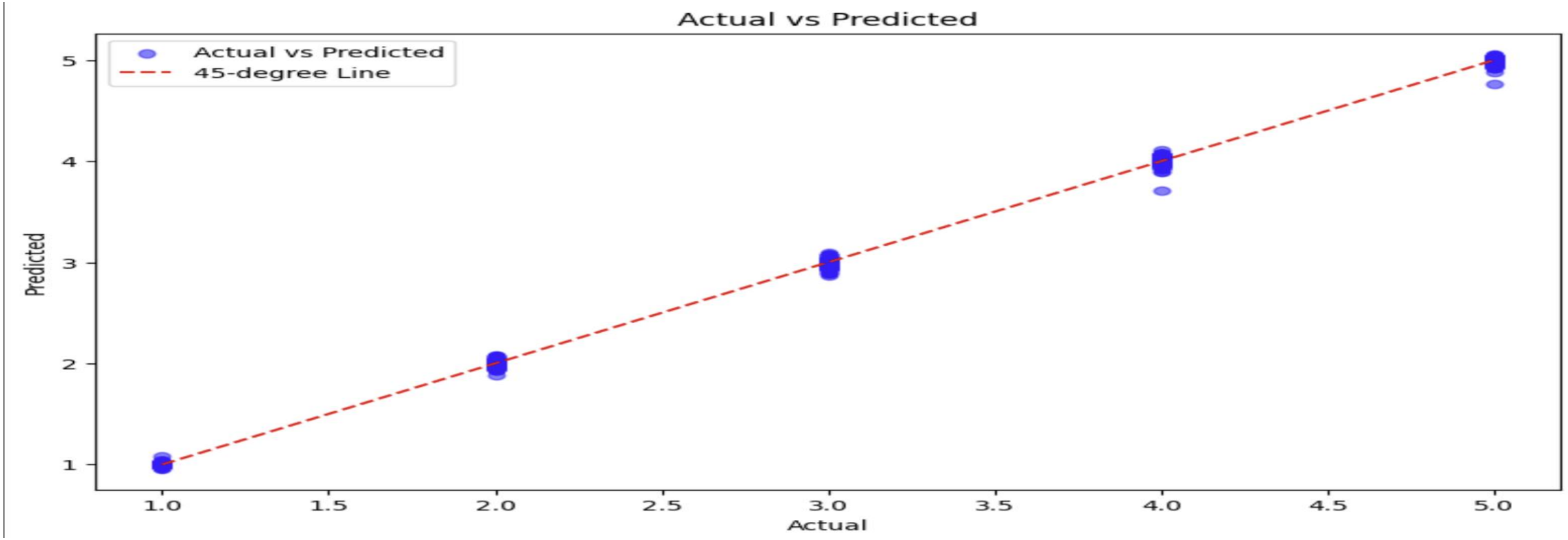
Confusion Matrices

Quantity sold with respect to avg discount

Modeling Methods

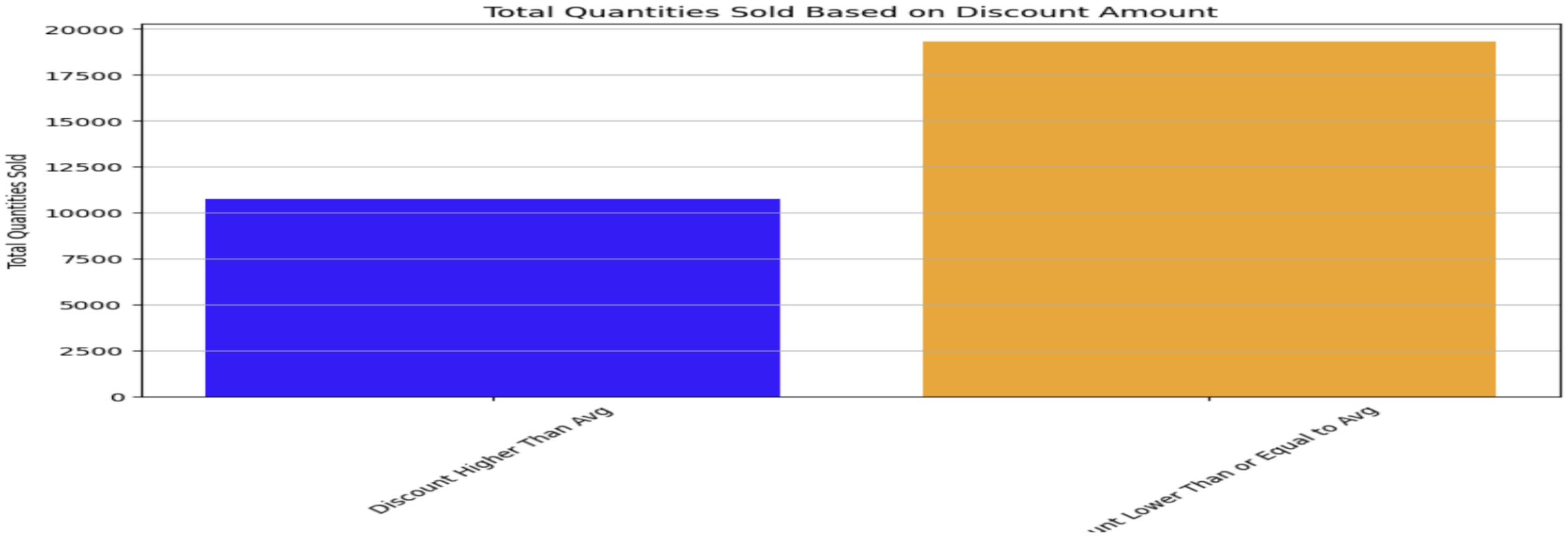
Modeling Methods

- Quantity Forecasting :- We've built a Regression model (CatboostRegressor) to predict quantity sales, because this will help business to understand what will be the future market demand.
  The better we predict this, the better we can make business action, like 'stock our shelves' or managing inventory according to the forecasted number.



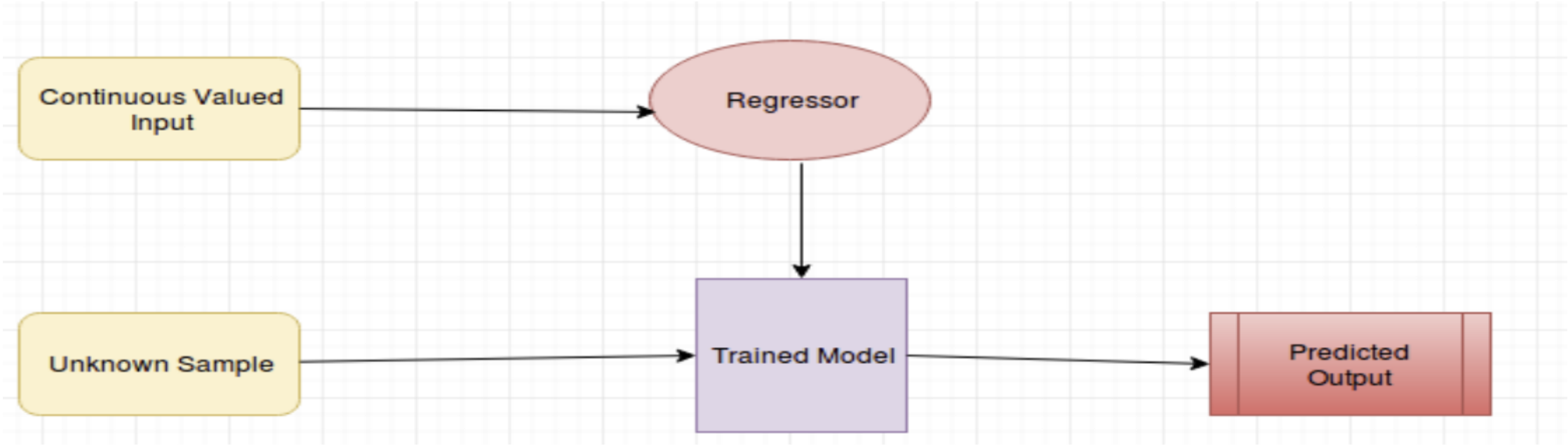
What Factors into Our Predictions

- Impacting Feature :-The most important features which are creating the impact on our analysis are total price and Discount amount this means that the higher the unit price and higher the discount higher the quantities are purchased



Model Type and Rationale

How We Make Our Predictions :- Think our model like a weather forecast - It takes the current condition and predict the future outcome. We have used “Catboostregressor” which uses the historical data and predict the future outcome.

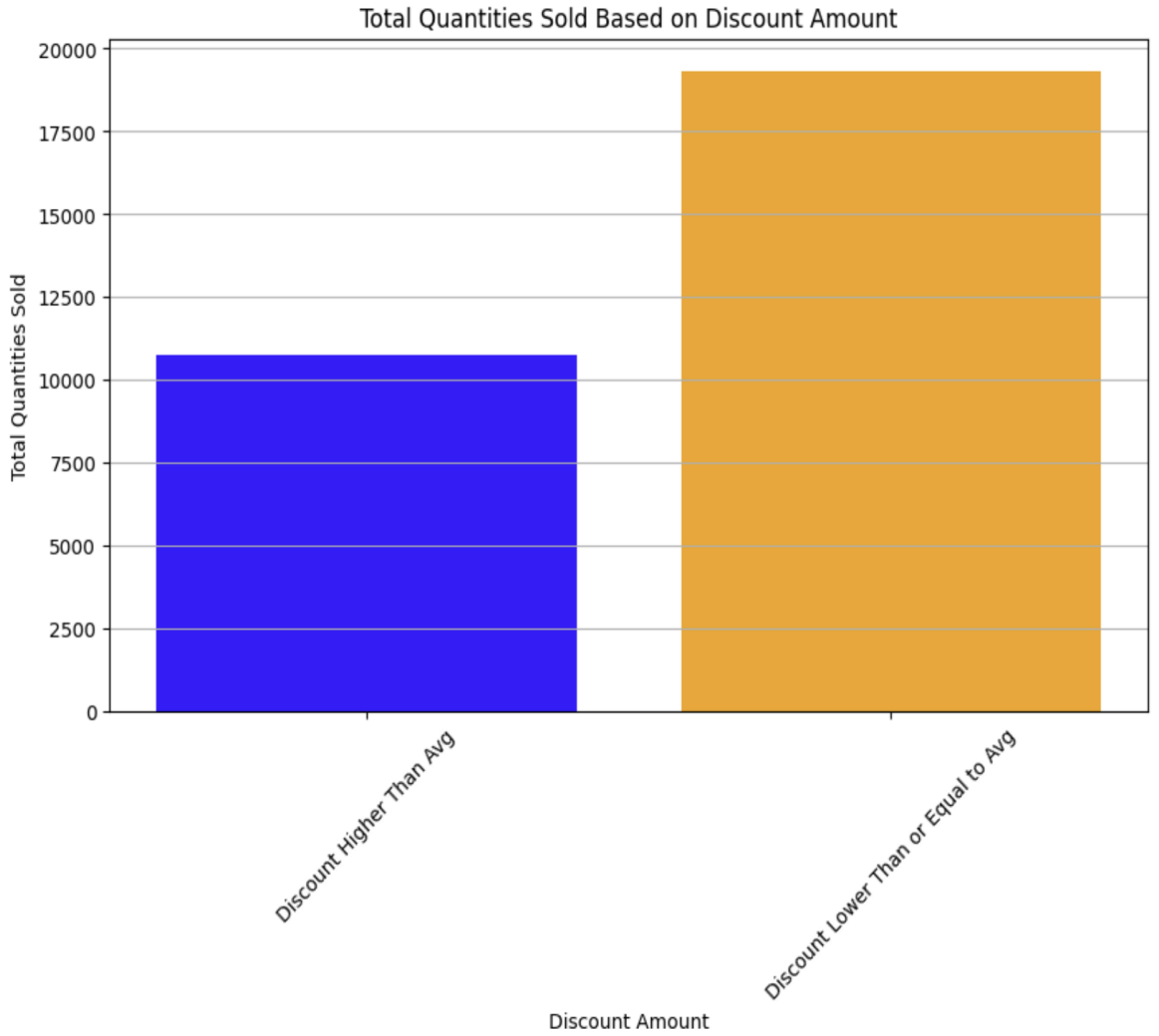


Findings

Insights that Drive Business Value

The most Important Insights we found during our analysis are

- Higher the discount and unit price of product the chances of customer buying that item is higher. which means our marketing team can set the prices accordingly leading the higher sale in quantity.



Recommendations and Next Steps

Next Steps for Further Analysis

- Our model's predictions are very close to real-world numbers, with an accuracy of 99.8%This means that for every 100 time model predict it will be right for 99% of time according to our Rsquared matrices but there is possibility that our model is overfitting / there is the data leakage which led to this results.
- To through with our developed model we should do further analysis by collecting more data / comparing with the other model which can provide us realistic numbers of our accuracy

Additional Information

- https://github.com/tirthmagnus/Practical-datascience

Thank You